

*B*IG DATA y sus APLICACIONES

MIGUEL SALDARRIAGA*, GERSON CORNEJO**
Y KEVIN GERSHY-DAMET***

El artículo investiga el surgimiento de la *Big Data* -base de datos de gran dimensión- y su actual importancia en el análisis económico. Las posibilidades de explotar estas fuentes de información son amplias y permiten formular hipótesis que antes no podían abordarse. Además, la *Big Data* representa una herramienta muy útil para los diseñadores de política.



* Jefe, Departamento de Indicadores de la Actividad Económica del BCRP.
miguel.saldarriaga@bcrp.gob.pe



** Especialista, Departamento de Indicadores de la Actividad Económica del BCRP.
gerson.cornejo@bcrp.gob.pe



*** Especialista, Departamento de Indicadores de la Actividad Económica del BCRP.
Kevin.gershydamet@bcrp.gob.pe

1. ¿QUÉ ES BIG DATA?

El análisis empírico de las bases de datos de gran dimensión es conocida con el nombre de *Big Data* y su aparición ha revolucionado tanto el análisis económico como el manejo estadístico. Algunas de sus características son la disponibilidad en tiempo real, la gran escala, la información sobre nuevas variables y la mayor complejidad en cuanto a estructura.

En primer lugar, la gran revolución inducida por esta información concierne a los medios de interacción estadística, que pueden ser aplicativos diseñados para búsquedas en internet o el uso de servicios gubernamentales, ambos en tiempo real. El hecho de tener información en tiempo real genera la posibilidad de automatizar algunos procesos, así como de mejorar la capacidad predictiva sobre múltiples variables de interés.

En segundo lugar, las nuevas bases de datos están compuestas por millones de observaciones y numerosas variables, las cuales brindan mayor representatividad sobre los resultados estadísticos que puedan ser calculados. En contraste, hasta la década anterior, los modelos contaban con una limitada capacidad predictiva debido a la escasez del número de observaciones y la incapacidad de medir algunas variables.

Tercero, gran parte de la información que se obtiene actualmente se encuentra vinculada a actividades económicas que anteriormente no era factible monitorear. Un ejemplo claro son las bases de datos generadas a partir de las redes sociales, las cuales son de interés para los investigadores sociales y conexiones interpersonales.

Cuarto, esta información posee estructuras de mayor complejidad y dimensionalidad, ya que proviene del monitoreo de actividades que se desarrollan secuencialmente, como el historial de compras de los consumidores. Esta complejidad representa un desafío para la teoría económica y el análisis econométrico, pues se generan múltiples opciones para el tratamiento del comportamiento individual de los agentes económicos.

El uso de la *Big Data* es amplio y puede ser aplicado a múltiples ciencias. Algunos de los campos que han aprovechado sus beneficios son el análisis de la dinámica poblacional, el cambio social, y los riesgos en la salud pública, además de muchas áreas de la ciencia económica como el desarrollo y crecimiento. Sin embargo, la disciplina con mayores ganancias gracias a estas bases de datos son aquellas dedicadas al estudio del comportamiento humano.

2. BIG DATA Y ECONOMÍA

Con respecto a años previos, la granularidad de la información con la que se cuenta actualmente permite responder preguntas que antes no podían abordarse; pero también

GRÁFICO 1 Principales características del Big Data



FUENTE: HIGHPOINT SOLUTIONS.
ELABORACIÓN: PROPIA.

brinda una oportunidad para comenzar a plantear hipótesis acerca de lo que sucede en tiempo real (proceso conocido como “*nowcasting*”). De hecho, gracias a la *Big Data* se cuenta con datos que antes no podían ser medidos o que eran difíciles de controlar. Por ejemplo, la georreferenciación individual ya no es un problema y las transacciones realizadas dentro del país pueden seguirse en tiempo real para cada individuo.

Estas novedades posibilitan dar un mayor desarrollo empírico a varias ramas de la economía. Un ejemplo claro es la economía del comportamiento: los historiales de tarjetas de crédito o débito otorgan acceso a las compras de cada persona, y dicha información permite realizar estudios sobre la toma de decisiones del consumidor, sus reacciones ante las ofertas por producto, sus patrones de endeudamiento y su exposición a la publicidad. Además, pensando de manera más agregada, surge la alternativa de generar estudios demográficos más detallados que se basen en patrones de consumo, tendencias de compra y de cuidado de la salud.

Por otro lado, a nivel de empresas, los registros de ventas permiten acceder al historial de transacciones realizadas en cada minuto del día. En ese sentido, la coordinación con empresas privadas puede ser clave para realizar estudios detallados sobre la evolución de las ventas por sector. En esta línea, el crecimiento de la capacidad de registro de la *world wide web* brinda una oportunidad para realizar un análisis más profundo sobre las tendencias de consumo mencionadas anteriormente. Por ejemplo, las medidas usadas con frecuencia para medir la efectividad de las publicidades virtuales son la cantidad de *clicks* dados hacia un anuncio específico y la cantidad de búsquedas realizadas por cada tema (Einav, L. y J. Levin, 2014).

Por ello, con la *Big Data* es posible generar experimentos conductuales que, a diferencia de los diseños experimentales realizados anteriormente, poseen mayor tamaño y alcance, lo cual puede brindar una mayor robustez a los resultados obtenidos. Además, para los diseñadores de política es posible mejorar la medición de variables de interés como el consumo y el empleo. Por ejemplo, a partir de los reportes en tiempo real de la plataforma “SpendingPulse” de MasterCard es posible obtener medidas sobre la actividad comercial y el gasto en consumo con bastante anticipación a las estadísticas oficiales. En la misma línea, con la información recolectada por los servicios de Medicare, es posible obtener información clínica de los pacientes a bajo costo y con rapidez debido a los servicios electrónicos.

En el mismo sentido, es posible mejorar las predicciones y el seguimiento de una serie de variables macroeconómicas y construir estadísticas propias a partir de las noticias reportadas por los periódicos. Este tipo de trabajos no solo requiere una gran capacidad computacional para el procesamiento de los datos, sino que también demanda mucha habilidad para el ordenamiento de los mismos, debido a su alta dimensionalidad.

Queda claro que gracias a la *Big Data* puede darse un seguimiento de mayor calidad a múltiples ámbitos de la economía. Las posibilidades de explotación de estas fuentes de información son vastas y dan pie a formular hipótesis que antes no podían abordarse; además representa una herramienta muy útil para los diseñadores de política.

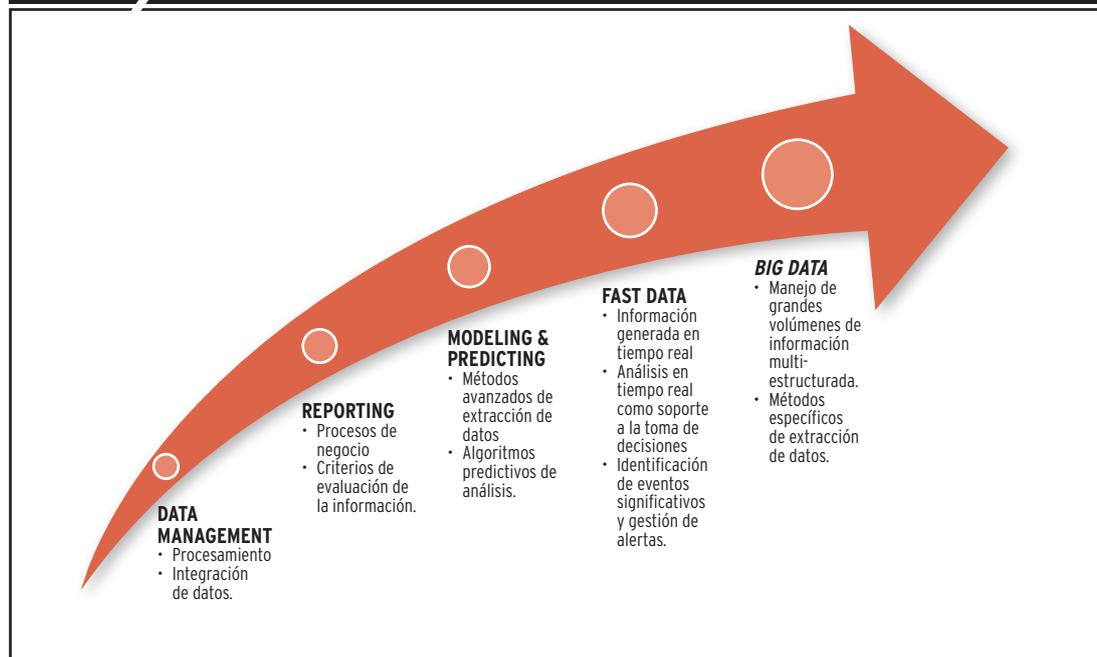
3. BIG DATA PARA LA TOMA DE DECISIONES

El uso creciente de la *Big Data* ha estado motivado por el sector privado como medio para mejorar la eficiencia de sus proyecciones y observar patrones de comportamiento, todo ello con la finalidad de aumentar su rentabilidad y superar a sus competidores. En el caso del sector público, por su parte, existen múltiples experiencias exitosas que confirman que existen altos beneficios para los diseñadores de política (cabe resaltar que el uso de *Big Data* en el ámbito estatal no implica que la fuente de los datos deba ser generada por el sector público). Al respecto, las principales ventajas en términos de toma de decisiones son que permite (i) profundizar el conocimiento sobre un tema y (ii) mejorar la capacidad predictiva.

De acuerdo con el Banco Mundial (2014) existen dos procesos que se pueden usar para “fortalecer el entendimiento de sistemas complejos con el fin de permitir un mejor proceso de diseño de política”: (i) utilizar la *Big Data* para analizar comportamientos fuera de la instancia del gobierno con el objetivo de mejorar el proceso de toma de decisiones; y (ii) utilizarla para analizar el comportamiento al interior de una institución, con el objetivo de mejorar un proceso interno.

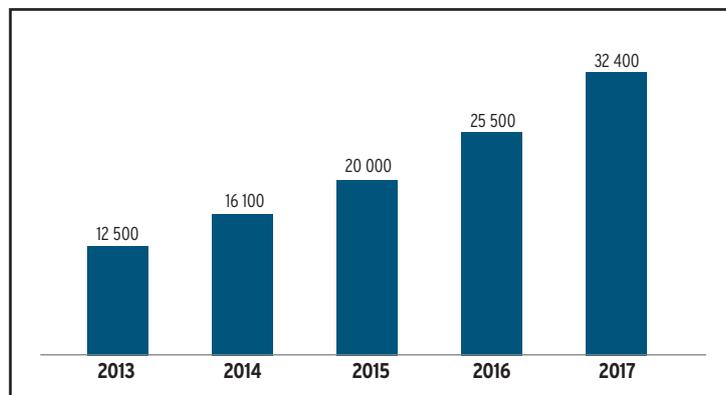
Respecto al primer punto, por ejemplo, ante la ausencia de información estadística de calidad, para algunos países de África se han realizado estimados de la actividad económica usando la emisión de luces vía satélite. En Indonesia, un seguimiento a los *tweets* que mencionaban las

GRÁFICO 2 ■ Evolución del manejo de información



FUENTE: HIGHPOINT SOLUTIONS (ELABORACIÓN PROPIA).

GRÁFICO 3 ■ Previsión de crecimiento del mercado *Big Data* (Tecnología+servicios) (En millones de US\$)



FUENTE: INTERNATIONAL DATA CORPORATION.
ELABORACIÓN: PROPIA.

variaciones en el precio del arroz permitía monitorear la inflación de precios de alimentos. Finalmente, en Irlanda se probó que los comentarios en foros y en diarios que muestran frustración y confusión adelantaban en tres meses las estadísticas de desempleo. Con relación al segundo punto, por ejemplo, el gobierno sueco usó información de sus usuarios para enviar mensajes de texto personalizados con el fin de mejorar la rapidez en la presentación de los formularios de declaración de impuestos.

En general, la *Big Data* permite obtener un mayor nivel de granularidad y detalle, que los diseñadores de política pueden considerar para tener resultados más precisos sobre determinadas políticas para grupos distintos de la población. Muchas de estas bases de datos se puedan obtener en tiempo real y permiten responder de manera inmediata. Por ejemplo, tras un desastre natural es posible conseguir información sobre poblados a través de los teléfonos celulares e imágenes satelitales, lo que a su vez acorta el tiempo de respuesta, mejora la eficacia de la misma y facilita la retroalimentación en tiempo real con el objetivo de diseñar e implementar mejores programas y políticas públicas.

Otra mejoría del uso de *Big Data* es que permitirá validar de manera más precisa los resultados de cualquier experimento que se desee plantear. Tal como señalan Einav y Levin (2014), al contar con mayor información se podrán realizar mejores mediciones de los efectos y resultados

de cualquier política. También se podrían probar distintas estrategias para medir la robustez de los resultados y la validez de los principales supuestos.

Finalmente, el uso de *Big Data* en la elaboración de propuestas de políticas públicas de alcance internacional ha sido ampliamente reconocido. Por ejemplo, en 2014 el Banco Mundial lanzó la *Big Data Innovation Challenge* con el objetivo de financiar 14 proyectos que usaran *Big Data* como su fuente principal de información. La diversidad generada da una idea del abanico de posibilidades que puede haber: mejorar los derechos de propiedad usando datos geoespaciales, medir la pobreza vía satélite, entender cómo la distribución de la infraestructura afecta al crimen, medir las condiciones de la carretera usando información del celular, entre otros.

4 PRINCIPALES DESAFÍOS

Algunos de los principales retos que enfrenta el uso de *Big Data* es determinar qué tipo de información se puede extraer y analizar, y cómo “hacer la pregunta correcta”, como señalan Einav y Levin (2014). La tecnología por sí sola no es suficiente para entender e interpretar los resultados; por eso, el estudio debe considerar aspectos como el sesgo de selección, la representatividad de la muestra y la forma en que se generaron los datos.

El exceso de información puede ser engañoso: si bien se pueden detectar muchas correlaciones, no siempre es posible saber cuáles tienen valor económico. Además, por la disponibilidad que la *Big Data* trae, ésta tiene limitaciones como la capacidad computacional y las habilidades informáticas necesarias para almacenar y organizar los datos. Esta última se asocia al cuidado que se debe tener al realizar proyecciones con mayor rigurosidad estadística, porque la gran cantidad de información hace que la estructura de los datos sea más compleja que lo usual (corte transversal, series de tiempo o datos de panel).

Finalmente, existen retos relacionados al acceso y la preparación de los datos. En cuanto al primero, la mayor parte de la información pertenece a instituciones privadas o públicas en las cuales los datos deben ser generados o no pueden ser compartidos de manera directa. Por otro lado, muchas veces generar directamente estos datos puede ser muy costoso.

BIBLIOGRAFÍA

- Einav, L. y J. Levin, (2014). “Economics in the age of big data”. *Science*, 346 - 6210, p. 715-721.
- Einav, L. y J. Levin, (2014). “The Data Revolution and Economic Analysis”. *Innovation Policy and the Economy*, Vol. 14.
- Imbens, G y Abadie, A. (2011), “Bias-Corrected Matching Estimators for Average Treatment Effects”. *American Statistical Association, Journal of Business and Economic Statistics*, Vol. 29 N°1.
- The World Bank Group. (2014). “Big Data in Action for Development”. Washington, DC. 68 pp.
- The World Bank Group, (2016). “Big Data Innovation Challenge: Pioneering Approaches to Data-Driven Development”. Washington, DC. 88 pp.